Topological and semantic Web based method for analyzing TGF-β signaling pathways

Jean Coquet, Geoffroy Andrieux, Jacques Nicolas, Olivier Dameron, Nathalie Théret

ANALYSIS OF TGF-β SIGNALIZATION PATHWAYS THANKS TO TOPOLOGICAL AND SEMANTIC WEB METHOD

Jean COQUET [1,2], Geoffroy ANDRIEUX [3], Jacques NICOLAS [1], Olivier DAMERON [1] and Nathalie THERET [2]

[1] Université de Rennes 1 - IRISA/INRIA, UMR6074, 263 avenue du Général Leclerc, 35042 Rennes, Cedex, France
[2] Université de Rennes 1 - IRSET EA 4427 SeRAIC, IFR140, 2 avenue Pr Léon Bernard, Rennes, Cedex, France
[3] German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

CONTEXT

- TGF-β = Transforming Growth Factor β
  - necessary in healthy situation
  - deleterious role in fibrosis and cancer
- Pleiotropic effects of TGF-β are linked to the complex nature of its activation and signaling networks

BIOLOGICAL QUESTIONS:
Which pathways are cancer-specific?
Which mechanism is involved in the healthy → pathological transition?

SIGNALIZATION MODEL

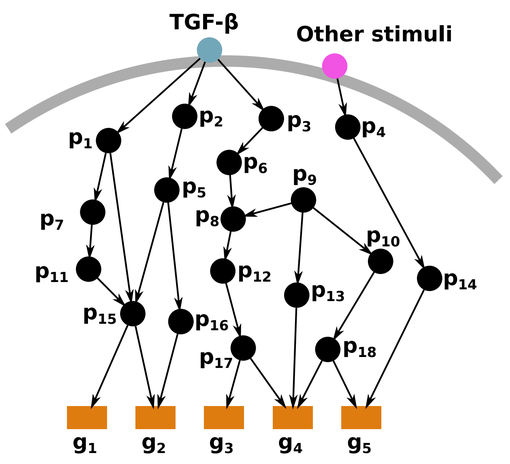
- Model of TGF-β signal propagation based on guarded transitions [Andrieux et al, 2014] represented in CADBIOM.
- Pathway Interaction Database (PID) transform into a single unified model of signal transduction

DATA ANALYSIS PROBLEM:
16,000 chains of reactions linking TGF-β to at least one of 159 target genes.

How to explore the combinatorial complexity of cell signaling?

ANALYSIS OF SIGNALING PATHWAYS : STEPS METHOD

I

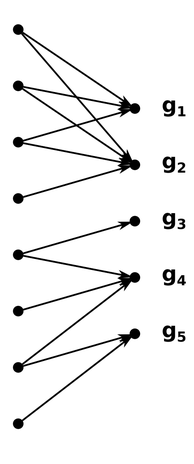


RAW DATA

- *TGF-β signalizations influence some genes*
- *Other stimuli in addition to TGF-β*

II

$s_1 = \{p_1, p_7, p_{11}, p_{15}\}$
$s_2 = \{p_1, p_{15}\}$
$s_3 = \{p_2, p_5, p_{15}\}$
$s_4 = \{p_2, p_5, p_{16}\}$
$s_5 = \{p_3, p_6, p_8, p_{12}, p_{17}\}$
$s_6 = \{p_3, p_6, p_8, p_9, p_{13}\}$
$s_7 = \{p_3, p_6, p_8, p_9, p_{10}, p_{18}\}$
$s_8 = \{p_4, p_{14}\}$



GENES INFLUENCED BY PROTEIN SETS

- *Signaling pathway = proteins set*
- *Proteins set influence at least one gene*

III

Protein sets

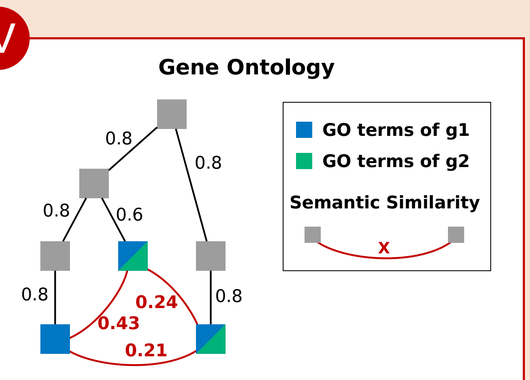
Genes	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8
g_1	1	1	1	0	0	0	0	0
g_2	1	1	1	1	0	0	0	0
g_3	0	0	0	0	1	0	0	0
g_4	0	0	0	0	1	1	1	0
g_5	0	0	0	0	0	0	1	1

Formal Concepts
(1) $\{g_1, g_2\} \times \{s_1, s_2, s_3\}$
(2) $\{g_2\} \times \{s_1, s_2, s_3, s_4\}$
(3) $\{g_3, g_4\} \times \{s_5\}$
(4) $\{g_4\} \times \{s_5, s_6, s_7\}$
(5) $\{g_4, g_5\} \times \{s_7\}$
(6) $\{g_5\} \times \{s_7, s_8\}$

CLUSTERS OF SIMILAR GENES AND PROTEIN SETS

- *Identify clusters of similar genes and protein sets with formal concept analysis*
- *Binary matrix*
- *Formal concepts = maximal bi-cliques*

IV



zScore for each genes group:

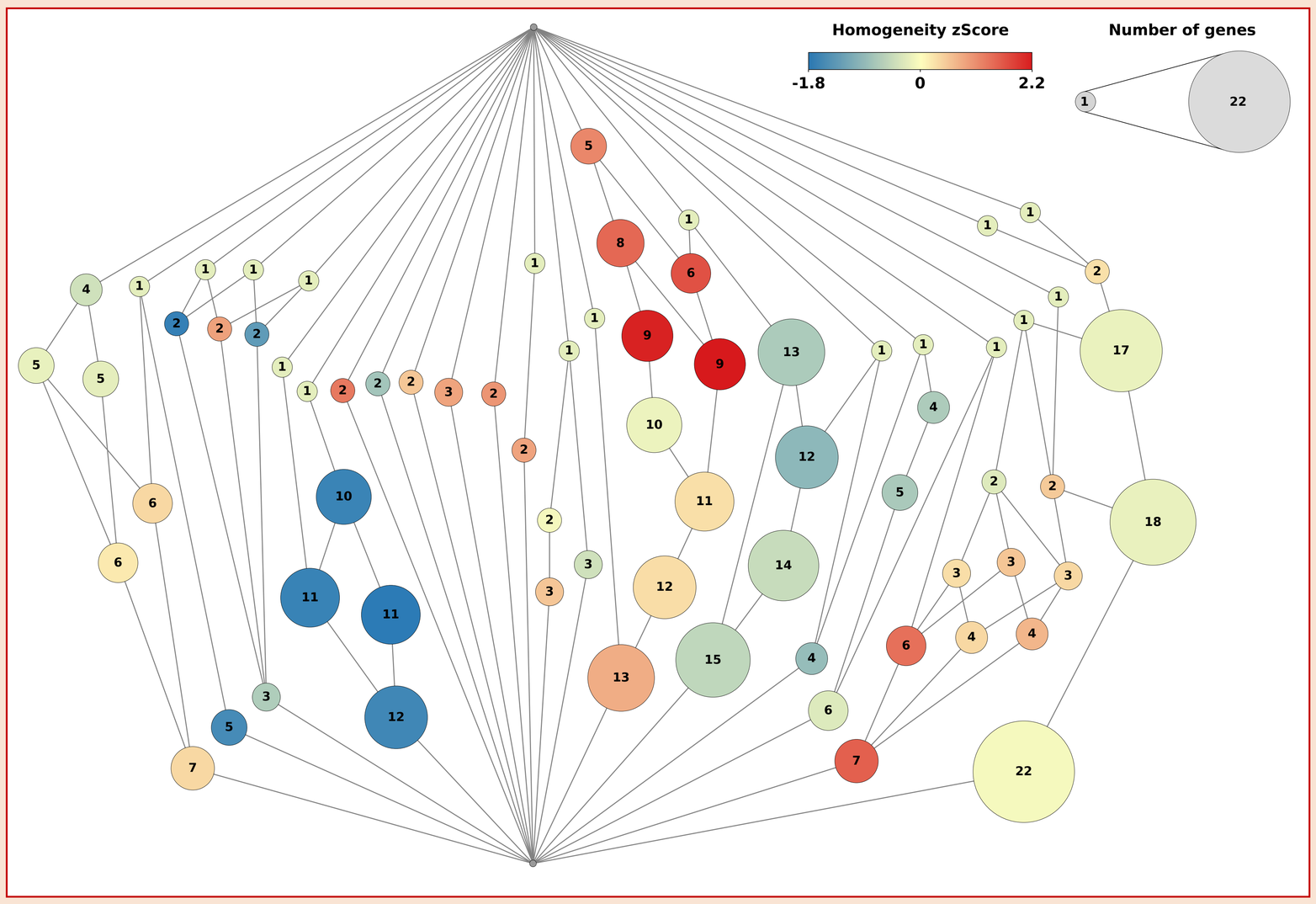
$$zScore(fc_i) = \frac{sim(fc_i) - m}{\sigma}$$

- **zScore > 1.96 : strong similarity**
- **zScore < -1.96 : low similarity**

ASSESS CLUSTERS' BIOLOGICAL RELEVANCE WITH SEMANTIC HOMOGENEITY

- *Formal concept homogeneity = zScore of similarity gene group*
- *Semantic similarity based on Wang measure [Wang et al, 2007]*

RESULTS



Concept lattice. Nodes are formal concepts, two nodes are linked if one is included into the other.

DISCUSSION

- Some formal concepts are highly connected
- Gene clusters can have:
  - High homogeneity = genes similar influenced by same pathways
  - Low homogeneity = genes dissimilar influenced by same pathways
- Chains of gene clusters included into one-another indicates that some influence sets are specific to some gene subsets

IN PROGRESS

- Validate the method on small synthetic models
  - Improve the method to find better gene clusters
  - Integrate the overlap between proteins sets
  - Try different methods of clustering
- Compute the semantic homogeneity of protein sets
- Take into account the structure underlying gene sets
- Compute the stability of formal concepts

CONCLUSION

Model of TGF-β signal propagation → **New method to analyze pathways: 1) Topological analysis 2) Clustering 3) Biological relevance** → Gene clusters influenced by same pathways with a biological relevance

References:
[Andrieux et al, 2014] Andrieux, G., Le Borgne, M., & Théret, N. (2014). An integrative modeling framework reveals plasticity of TGF-beta signaling. BMC systems biology, 8(1), 30.
[Wang et al, 2007] Wang, J. Z., Du, Z., Payattakool, R., Philip, S. Y., & Chen, C. F. (2007). A new method to measure the semantic similarity of GO terms. Bioinformatics, 23(10), 1274-1281.